

# FACE IDENTIFICATION USING MULTIPLE COMBINATION STRATEGY FOR HUMAN ROBOT INTERACTION

Do-Hyung Kim<sup>\*</sup>, Jae-Yeon Lee<sup>\*</sup>, Eui-Young Cha<sup>\*\*</sup>, Young-Jo Cho<sup>\*</sup>

*\* Intelligent Robot Research Division  
Electronics and Telecommunications Research Institute  
{dhkim008, leeji, youngjo}@etri.re.kr*

*\*\* Neural network and real world application laboratory  
Pusan Nat'l University.  
eycha@harmony.cs.pusan.ac.kr*

Abstract: The ability to recognize people is positively necessary for interaction between robots and people. In this paper, we introduce interesting security system that is able to recognize, track and remember user's face. System uses moving camera that has pan/tilt and zoom function like robot's eyes. In order to construct system, we propose the face identification method based on combining multiple features and applying multiple matching algorithms. Also, we address the benefits of multiple instances and show that identification ability can be improved by using proper combination strategy. *Copyright © 2005 IFAC*

Keywords: Pattern identification, Visual pattern recognition, Robot vision, Human-machine interface, Interaction mechanisms.

## 1. INTRODUCTION

For robot to communicate and cooperate with people, the ability to identify people is essential. This ability allows the robot to authenticate user and provide separate services for each person. In order to identify people, several identification technologies can be combined, such as face recognition, iris recognition, voice recognition, and semi-biometrics features such as height and volume that are helpful to distinguish individuals included certain target group. Among these identification technologies, the face identification technology is most crucial and widely used since it is more superior and natural than other technologies.

Numerous studies on face identification have been explored over the years. They can be divided into two broad categories, techniques based on template matching and techniques based on the computation of geometrical features (Brunelli and Poggio, 1993). The major methods for representation and recognition of human face are principal component analysis(PCA) (Turk and Pentland, 1991), elastic bunch graph matching (Wiskott, *et al.*, 1996), linear discriminant analysis(LDA) (Etemad and Chellappa, 1997), local feature analysis(LFA) (Penev and Atick, 1996), and so on. There are many reports that above-mentioned face recognition methods show excellent performances in restricted environment.

But, it seems difficult to deal with the variability in appearance due to changes in lighting condition, pose, and expression by using an only single method. From this point of view, recently, various researches have been tried to solve these problems by combining

---

<sup>+</sup> This work is a result of URC project sponsored by MIC of Korean government.

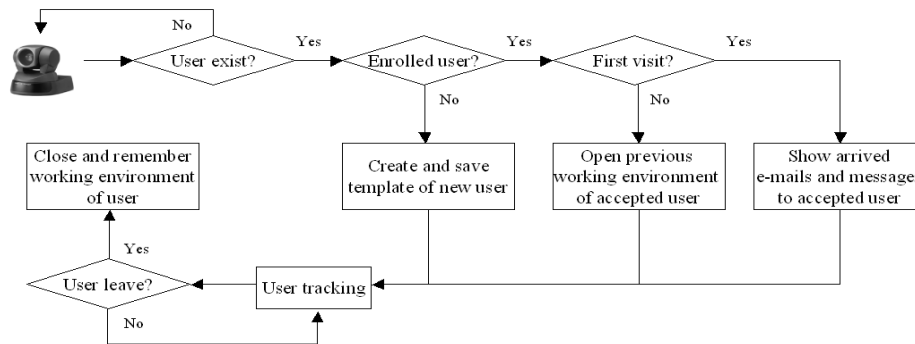


Fig. 1. The process sequence of security system.

mutually complementary features and classifiers. That is combining global and local feature (Yuchun, *et al.*, 2002), combining classifiers and several algorithms (Jie, 2002), and applying multiple instances and matching algorithms (Kittler, *et al.*, 1997). It has been reported that these approaches perform well in many real-world situation.

For robot to identify individuals, we propose the method based on combining multiple features and applying multiple matching algorithms. We describe the feature extraction approaches based on multiple principal component analysis and edge distribution. These features are projected on a new intra-person/extra-person similarity space that consists of several simple similarity measures, and are finally evaluated by a support vector machine. Also, we propose the methodology for fusing multiple instances of biometric data. In the experiments on a realistic database, our approach shows very encouraging result. We think that the proposed scheme is one of the essential considerations for building a real world face identification system.

Using the proposed algorithm, we construct the interesting face identification system for the security of multi-user working environment on PC. System is able to recognize, track, and remember user by using moving camera that has pan/tilt and zoom function like robot's eye. By providing separate working environment for each user, system can protect personal digital documents and information on PC.

This paper is organized as follows. In Section 2, the overview of our security system is described. Section 3 deals with an implementation. The detailed description of the proposed features is given and novel similarity space and methodology for fusing multiple instances are described. An evaluation model and the experimental results are shown in Section 4. Finally, Section 5 concludes this paper.

## 2. OVERVIEW OF SECURITY SYSTEM

Our security system has a moving camera on the top of the monitor. The moving camera Sony EVI-D100 has high-speed pan/tilt ability and uses a 10x optical zoom lens. This pan/tilt and digital zoom function

allow system to track and recognize user to more wide range. We can operate the camera by connecting to a system with a VISCA cable (RS-232C).

The simple process sequence of our security system is as follows. When a user who wants to use PC comes in front of the moving camera, system starts to search user's eyes in order to recognize user. After system searches a face region of user successfully, recognition process is started. If recognition result says that current user is already enrolled user, system allows accepted user to work on PC. In case of user's first visit to PC, system shows arrived e-mails and messages to accepted user. If current user is turned out unknown user, system enrolls new template of unknown user and then enrolled user is able to use PC. While accepted user works on PC, system with moving camera continuously tracks user's eyes using domain knowledge of eyes and motion information in order to know whether user leaves PC. When user stops working and leaves PC, system closes all windows that user operates for protecting personal digital documents and information, and remembers user and corresponding working environment information for user's next visit. For other users, above process sequence is applied. If previous visitor returns to work on PC again, system recognizes user and opens previous all windows of accepted user.

## 3. IMPLEMENTATION

In this section the detailed description of the proposed algorithms is described.

### 3.1 Face Detector

We used a face detection system developed by (Yoon, *et al.*, 2004). The face detector outputs the location of eyes. This face detector is able to detect eye components from face images under size variation, complex background and skew angle variations. Detector has two steps for the eye component detection algorithm. The first step is an eye-based edge verification algorithm that analysis the isolated region, which acquired by adaptive Sobel edge detector. The second step is a lip-based edge verification algorithm and masking-based eye

centroid detection algorithm when input frame have not isolated eye components.

### 3.2 Preprocessing

In order to establish correspondence between face images to be compared, each face image is preprocessed. This step is divided into two processes, geometric normalization for adjusting the location of facial features and photometric normalization for improving the quality of the face image. In geometric normalization, based on manually localized eye positions, we rotate, size scale, and then crop the face region from the original image. In the photometric normalization, in order to remove the interference background, we put a mask on the cropped face image. In order to compensate for illumination, we apply histogram equalization on the face image except the mask region. As part of intensity normalization, we apply standardization, that is, we subtract the mean value of the face image from each intensity value and divide by standard deviation.

### 3.3 Feature Extraction

We extract the multiple principal component analysis vectors based on three face regions and the edge distribution vector of the face image. These features are projected on the new intra-person/extra-person similarity space that consists of several simple similarity measures, and are finally evaluated by the support vector machine.

#### Multiple PCA

Principal component analysis(PCA) is the optimal linear method for reducing the dimensionality of a data set while retaining as much variation as possible in the data set. Based on this central idea of PCA, the eigenface method is the most widely used method for representation and recognition of human faces (Turk and Pentland, 1991).

Among the face identification techniques based on the eigenface method, it is reported that the extension of the eigenface to facial features leads to an improvement in the recognition performance (Moghaddam and Pentland, 1997). This can be viewed as a modular representation of a face, where a coarse description of the whole face is augmented by additional details in terms of salient facial features. Though these additional facial features improve performance, in order to select the correct regions of features, the computational expense increases and detection errors occur.

Therefore, this paper proposes eigenUpper and eigenTzone as shown in Fig. 2. Since the proposed feature regions are selected just based on the eye positions, additional computational expense and a detection error are minimized.

eigenUpper: This model is for compensating the changes in expression. We know that among salient

facial features a mouth is most sensitive to expression instinctively. Therefore, just by removing a mouth region, we can compensate for the changes in expression such as smiling and talking.

eigenTzone: This model includes only the main facial features, eye and nose. eigenTzone is designed for compensating for illumination. So, the regions sensitive to illumination, such as forehead and cheek are excluded. Also, the T-zone of a face contains very important details for representing human faces.

#### Edge Distribution

We consider that edge information is another important factor for representing human faces. It is well known that a caricaturist represents the characteristics of a human being using only several lines. From the point of view, we have a consideration that the edge distribution on a human face becomes a good feature for face identification.

In our algorithms, the edge density of face region is normalized by the value of 25% edge density as shown in Fig. 3. The procedure of edge density normalization is as follows.

- Make the edge image using a Sobel edge operator.
- Make the distribution histogram of edge intensity.
- Determine the threshold( $\theta$ ) at 25% rank from the highest intensity.
- Select pixels that have a higher intensity than the threshold( $\theta$ ).

Then, this normalized binary edge image is divided into 10×10 size block and each block has the edge density value itself as the feature value. In our system, the size of cropped face region is 80×90, so the edge density vector consists of 72 elements. Our algorithm is very simple, and we can overcome the weakness of edge operation to illumination efficiently.

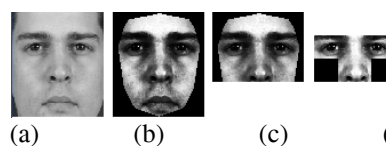


Fig. 2. The selected face region for multiple PCA : (a) cropped face region, (b) eigenface, (c) eigenUpper, and (d) eigenTzone.

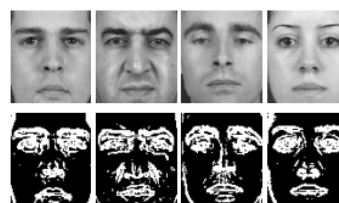


Fig. 3. The cropped image(top row) and the corresponding binary edge image normalized by the value of 25% edge density (bottom row).

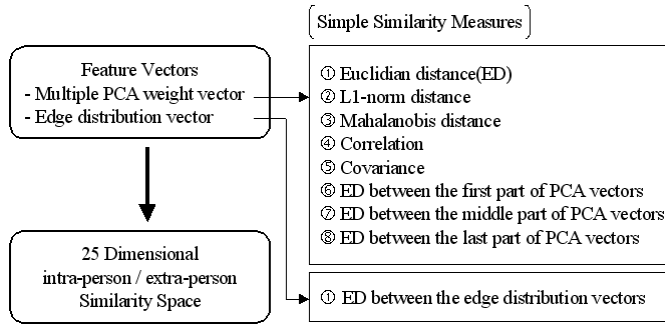


Fig. 4. The process sequence of our security system.

### 3.4 Identification

A Support Vector Machine(SVM) is basically a binary classifier based on the statistical learning model using the high dimensional virtual feature space (Burges, 1998).

In our algorithm, we classify the multiple PCA features and the edge distribution feature using a support vector machine. At this time, if these feature vectors are fed to SVM, one SVM per enrolled person trained by each person's features is necessary. A large number of SVMs causes the additional expense, the processing speed decrease and the memory increase in enroll process.

As the solution to this problem, we define a novel intra-person/extra-person similarity space that represents the correlation between the two feature vectors. The projection of the two feature vectors(one for the input face and the other for the enrolled face) on the similarity space is performed by several simple similarity measures.

They are :

- Euclidian distance(ED)
- L1-norm distance
- Mahalanobis distance
- Correlation
- Covariance
- ED between the first part of PCA vector
- ED between the middle part of PCA vector
- ED between the last part of PCA vector

Fig. 4. shows the architecture of the multiple matching algorithm designed in order to combine several features. By the above-mentioned similarity measures, the 8 similarities are created between the eigenface weight vector of the input face and that of the enrolled face. For eigenUpper and eigenTzone, the same process is adopted. Also one similarity value is created between two edge distribution vectors by Euclidian distance. By the combination of 3 similarity vectors(8 elements) and 1 similarity, finally similarity vector with 25 elements is created.

Since the similarity vector is created based on the similarity of two feature vectors, it is clustered into two classes in 25 dimensional similarity space. That

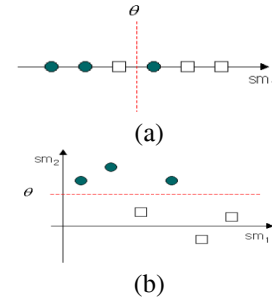


Fig. 5. Improvement of the capability of a classifier by expanding a similarity space

is, the intra-person similarity vector class and the extra-person similarity vector class. Each class is clustered around similar position regardless of person's characteristic and the number of persons in the database. And then SVM, the binary classifier, classify these classes efficiently.

As described, we try to expand the similarity space using several similarity measures. The capability of a classifier can be improved by expanding a similarity space, which can be explained with Fig. 5. As shown in Fig. 5, we have 2 classes to be classified. (a) shows the similarity space of one dimension consisting of one similarity measure( $sm_1$ ). It seems impossible to classify classes completely in the (a). Now, assume that we have another similarity measure( $sm_2$ ) that has a low correlation and a complementary relation with  $sm_1$ . If we expand the similarity space with new similarity measure, we can improve the capability of a classifier as (b).

We don't involve process to calculate a correlation between similarity measures and select a proper similarity measure. But, even if the similarity measure that has a high correlation with others is selected, it does not debilitate the classification capability of a SVM. So, we select the most general similarity measures optionally. But, in the future, the development and selection of the proper similarity measure is necessary for the effectiveness of the system. By defining a new similarity space, we can use just one SVM for classification, and alleviate the speed and memory problem.

## 4. EXPERIMENTAL RESULTS

### 4.1 Databases

For experiments, we use the Inha univ. database and the ETRI-A database. We capture images by a USB camera in general office environments and target the frontal face images without constraints on illumination and expression. The size of an image is  $320 \times 240$  pixels and the type is 24bit RGB color. The Inha univ. database and the ETRI-A database consist of 2,100 samples from 105 oriental persons and 1,120 samples from 56 oriental persons, respectively.

## 4.2 Standard Eigenface

In our evaluation model, the Inha univ. database is used just for making eigenfaces. We test only on the ETRI-A database. Therefore, persons who participate in making the eigenfaces don't exist in the database for the tests and eventually eigenfaces are always fixed independent of the database.

In a real-world system, the composition of the members in the database is always variable. If we don't have the fixed standard eigenfaces, we have to make the eigenfaces based on the current database. In this case, we have to remake them whenever persons are added to the current database. Actually, this frequent update of eigenfaces causes the useless computational expense and above all influences the feature vectors of the existing enrolled persons. For these reasons, by designing the above-mentioned evaluation model, we wish to do realistic and accurate experiments.

## 4.3 Experiments

In the ETRI-A database composed of 20 images per person, by considering the size of enroll data, enroll time, user convenience, performance, we empirically decide to use 5 images for enrollment and the remaining 15 images for test.

The similarity vector ( $S$ ) between the feature vector of the input image ( $If$ ) and the 5 feature vectors of enrolled images ( $Ef$ ) is created as follows :

$$S_k = \underset{i=1}{\overset{5}{\text{MAX}}}(M_k(If, Ef_i)) \quad (1)$$

Where  $i$  is the index of the enrolled image,  $k$  is the index of the similarity vector element ( $1 \leq k \leq 25$ ), and  $M$  is the similarity measure.

Table 1 Comparison of recognition rate for several feature vectors and similarity measures on the ETRI-A database

Feature Vectors	Similarity Measures	Recognition Rate (top ranked)
Eigenface Weight Vector	Euclidian	0.8381
	L1-norm	0.8702
	Mahalanobis	0.8848
	Correlation	0.8417
	Covariance	0.8429
Edge Distribution	Euclidian	0.7595
Eigenface Similarity Vector		0.9143
EigenUpper Similarity Vector		0.9095
EigenTzone Similarity Vector	SVM	0.8493
Multiple PCA Combination	Supervisor	0.9333
Edge Distribution & Multiple PCA Combination		0.9548

Table 1. shows the fraction of probes whose gallery match was top ranked for the several feature vectors and the similarity measures. As expected, the proposed algorithm outperforms eigenface weight feature vector by 0.07~0.12.

As a single similarity vector, the recognition rate of eigenUpper and eigenTzone is 0.9095 and 0.8493 respectively. Although the performance of these eigenfactors is a bit inferior to that of eigenface, we can improve the system performance by combining these eigenfactors. These results prove that the selected regions of eigenUpper and eigenTzone are suitable and give the efficient details for representing human faces.

As a single feature vector, the recognition rate of edge distribution feature vector is 0.7595, which is somewhat poor. But in combination with the multiple PCA features, the performance is improved remarkably. This result means that the proposed edge distribution feature is good enough to represent faces and the combination of the different feature extraction approaches (PCA texture-based approach / edge-based approach) is a good solution to the real-world problems.

Note that we use the fixed standard eigenfactors on the above experiments. The result shown in Table 1 is very encouraging and is a sufficiently practical level for many real-world applications.

## 4.4 Additional experiments

Above experimental results are for frontal mugshot-like images. But, for robot to identify people it is not to use just 1 frontal still image. Robot can capture consecutive images easily and identification ability can be improved by combining multiple similarities properly. So we investigate the benefits of multiple instances for ETRI-B database.

The ETRI-B database consists of 750 samples from 10 oriental persons. All samples from the Inha univ. database and ETRI-A database are still images. However, 75 samples per person from the ETRI-B database include 5 still images and 10 sets that consist of 7 consecutive images per set. The sets of consecutive images are constructed for multiple instances test and 5 still images are used for enrollment.

The number of consecutive images used an experiment is 1, 3, 5 and 7. As an identification algorithm, we select edge distribution & multiple PCA combination algorithm that shows the best performance in previous experiment. In order to combine the similarity value of each sample, we decide to select the following 4 combination strategies (Kittler, *et al.*, 1997).

**Table 2 Equal Error Rate (EER) for the number of consecutive images and combination strategies**

No. Images	Combination Strategies			
	Max	Min	Median	Average
1	3.15	3.15	3.15	3.15
3	1.33	5.06	1.78	1.78
5	0.78	6.22	1.17	1.39
7	0.39	5.00	0.94	0.94

- Selection of maximum similarity value (Max)
- Selection of minimum similarity value (Min)
- Selection of median similarity value (Median)
- Average of all similarity values (Average)

It is general to evaluate the recognition performance of system using cumulative recognition rate. However, from the experiments on just 1 still image of ETRI-B database, top-ranked recognition rate of 0.99 is achieved. This rate is too high to evaluate efficiently combination strategies for consecutive images. So we use the equal error rate instead of cumulative recognition rate. We may think that the lower the equal error rate, the better performance.

Table 2. shows the equal error rate(EER) for the number of consecutive images and combination strategies. As shown in the result, for the combination strategies except the Min, the performance is improved as increasing the number of consecutive images. Among these combination strategies, the Max is the best performer in this experiment. By selecting maximum similarity value, the Max can avoid errors due to changes in expression and pose. Though the maximum similarity value for impostor also is increased, the rate of increase for client is higher than that for impostor. Generally, the performance is improved as increasing the number of consecutive images. But, Note that the reduction of the performance almost saturates for 3 or 5 samples. Therefore when we build a real-world face identification system, by considering a trade-off between operating time and performance, selection of proper the number of test sample is necessary. We use 5 consecutive images for identification module in our system.

## 5. CONCLUSION

In this paper, we introduce efficient face identification system for robot to recognize and remember people. We propose feature extraction approaches based on multiple principal component analysis and edge distribution. These features are projected on the new intra-person/extra-person similarity space that consists of several simple similarity measures, and are finally evaluated by the support vector machine. Also, we address the benefits of multiple instances and show that identification ability can be improved by using proper combination strategy.

Using the proposed algorithm, we construct the interesting face identification application for the

security of multi-user working environment on PC. System is able to recognize, track, and remember user by using moving camera that has pan/tilt and zoom function like robot's eye. In addition to above-mentioned application, our face identification module is loaded on ETRO, our intelligent service robot.

So far, we have paid much attention on development of robust face recognition algorithms to solve the face problem caused by variability in appearance. The performance is evaluated by accuracy criteria such as equal error rate and cumulative recognition rate. For robot, however, the ability to consistently recognize and remember people who interact with robot is more important than recognition accuracy per image. Currently, we are researching face recognition architecture of robot and testing performance continuously with new criteria.

## REFERENCES

- Brunelli, R. and T. Poggio (1993) "Face recognition: features versus templates," *IEEE Transactions on PAMI*, vol. 15, no. 10, pp. 1042-1052
- Burges, C.J.C. (1998) "A tutorial on support vector machines for pattern recognition," *Data Mining & Knowledge Discovery*, vol. 2, no. 2, pp. 121-167.
- Etemad, K. and R. Chellappa (1997) "Discriminant Analysis for Recognition of Human Face Images," *J. Optical Society of America*, vol.14, pp.1724-1733.
- Jie Zhou (2002) "Face Recognition by Combining Several Algorithms," *Int'l Conf. Pattern Recognition*, vol.3, pp.497-500.
- Kittler, J., J. Matas, K. Johnson and M. U. Ramos Snchez (1997) "Combining evidence in personal identity verification systems," *Pattern Recognition Letters*, vol.18, pp.845-852.
- Moghaddam, B. and A. Pentland (1997) "Probabilistic visual learning for object representation," *IEEE Transactions on PAMI*, vol. 19, no. 7, pp. 696-710.
- Penev, P.S. and J.J. Atick (1996) "Local Feature Analysis : A General Statistical Theory for Object Representation," *Computation in Neural Systems*, vol.7, No.3, pp.477-500.
- Turk, M. and A. Pentland (1991) "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol.3, no.1, pp.71-86.
- Wiskott, L., J.M. Fellous, N. Kruger and C. Malsburg (1996) "Face Recognition by Elastic Graph Matching," *IEEE Transactions on PAMI*, vol.19, no.7, pp.775-779.
- Yoon, H.S., D. H. Kim and J.Y. Lee (2004) "Robust Eye Detection Method Using Domain Knowledge," *AIA-2004(Artificial Intelligence and Applications)*, pp. 559-563.
- Yuchun Fang, Tieniu Tan and Yunhong Wang (2002) "Fusion of Global and Local Features for Face Verification," *Int'l Conf. Pattern Recognition*, vol.2, pp.382-385.